Supplementary materials to the article: ‘Comparison of variance-based and moment-independent global sensitivity analysis approaches by application to the SWAT model’

Farkhondeh Khorashadi Zadeh, Jiri Nossent, Fanny Sarrazin, Francesca Pianosi, Ann van Griensven, Thorsten Wagener and Willy Bauwens

(2016, Environmental Modelling & Software)

Note that references cited in this Supplementary Materials are reported in the reference section of the article.

Section A. Anderson-Darling statistic as an alternative to the Kolmogorov-Smirnov statistic in the PAWN method

As an alternative to the Kolmogorov-Smirnov (KS) statistic, the Anderson-Darling (AD) (Anderson and Darling, 1952) statistic can be used to calculate the sensitivity index in the PAWN method. Similar to the KS statistic, the AD statistic is used to measure the difference between the unconditional and conditional CDFs of the model output, as below.

$$AD(X_i) = \frac{1}{N_u N_c} \sum_{j=1}^{N_T-1} \frac{(M_j N_T - N_c j)^2}{j(N_T - j)} \tag{A-1}$$

where N_u and N_c are the sizes of the unconditional and conditional samples, respectively, N_T is the sum of N_c and N_u, and M_j is defined as the number of conditional samples less than or equal to the jth smallest in the combined (unconditional and conditional) samples.

Similar to the KS statistic, since AD statistic depends on the conditioning value of X_i, a statistic (maximum in our case) over the n randomly sampled values for the fixed parameter X_i is used to calculated the PAWN sensitivity index, T_i^{AD}.

$$T_i^{AD} = \underset{X_i}{stat}[AD(X_i)] \tag{A-2}$$

Unlike the KS statistic, the AD value does not vary between 0 and 1. Therefore, to obtain sensitivity indices between 0 and 1, the sensitivity indices are normalized by dividing them by the sum of the sensitivity indices.

If the unconditional distribution $F_Y(Y)$ completely coincides with the conditional distribution $F_{Y|X_i}(Y)$, then the AD value is zero, indicating that X_i is non-influential. Clearly, a larger distance between two CDFs results in a higher value for the AD statistic, representing a higher influence for the given parameter (factor prioritization). Moreover, the two-sample Anderson-Darling (AD) test (Pettitt, 1976) can be applied in the PAWN method to identify the non-influential parameters with a given significance level *α* (typically set to 5%).

In this research, we investigated the effect of applying the AD statistic rather than the KS statistic in the application of the PAWN method to the SWAT model. The PAWN sensitivity indices and the parameter ranking results for the NSE and ME performance measures using the KS and the AD statistics are illustrated in Figure A.

For the NSE performance measure (Figures A(a) and A(b)), applying the AD statistic instead of KS results in different PAWN sensitivity indices for the SWAT parameters. However, interestingly, these two statistics provide very similar parameter ranking results, especially for the top ranked parameters. As stated in Section 3.3, the parameters ranked 12 and worse have the same PAWN sensitivity indices (close to zero), based on the KS value. According to the AD statistic, the sensitivity indices of this group of parameters (i.e. ranked 12 and worse) are also almost zero. Similarly, for the ME performance measure (Figures A(c) and A(d)), applying the KS and AD statistics leads to similar parameter rankings, especially for the most influential parameters. The top 11 ranked parameters, based on the KS and AD statistics, are exactly identical. Therefore, it is concluded that although the KS and AD statistics have a different definition and compare the CDFs in a different way, they provide similar parameter ranking results for the SWAT parameters in this study. This comparison increases the reliability of the conclusions drawn from the PAWN sensitivity analysis at almost no additional computing cost, since the different statistics are computed using the same samples and no additional model evaluations are required.

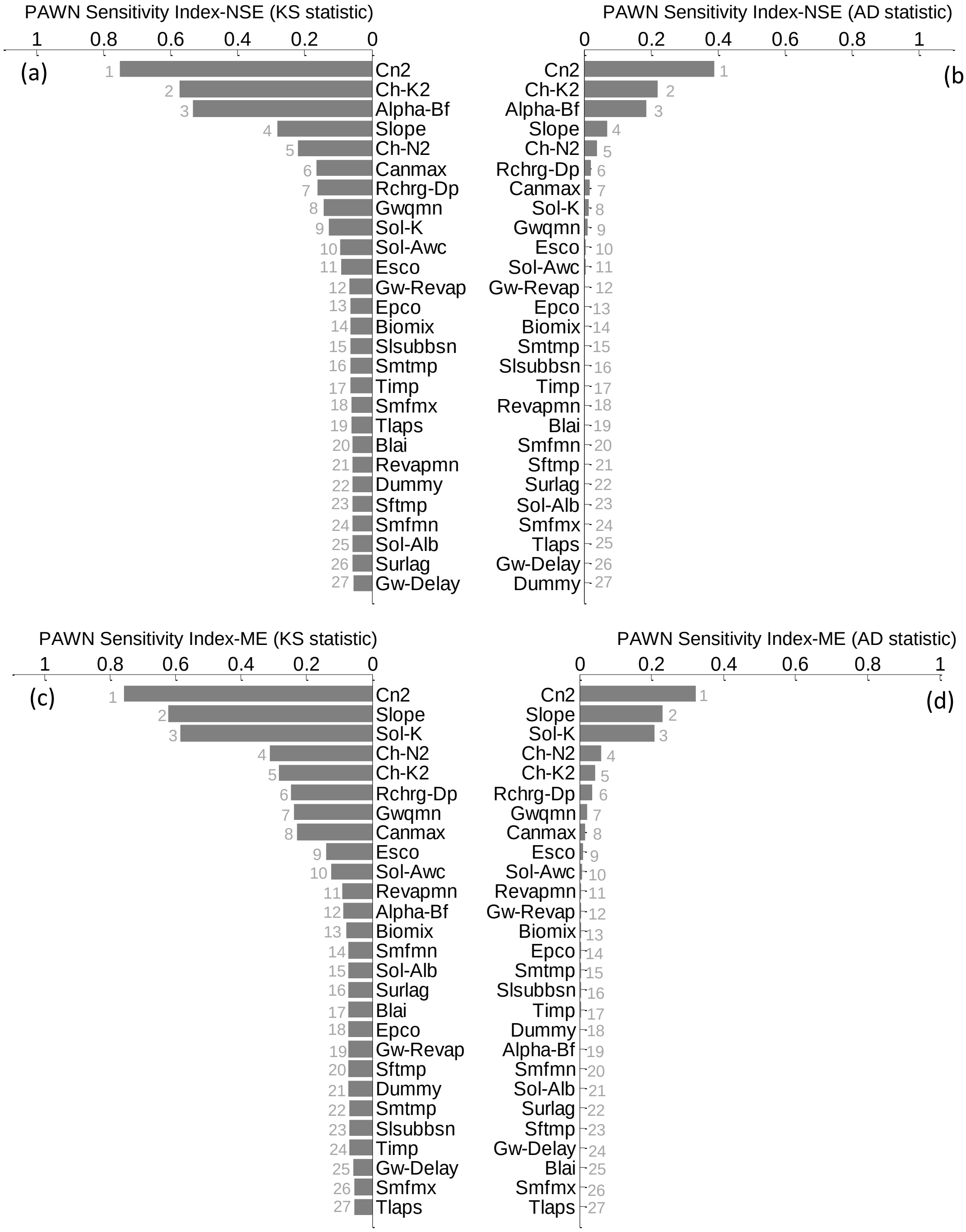

Figure A. Applying the KS and AD statistics provides almost the same parameter ranking results for the NSE ((a) and (b)) and the ME ((c) and (d)) performance measures in the PAWN method, especially for the top-ranked parameters.

Section B. Comparison of the bootstrapping results with and without replacement

To create resample data in the bootstrap technique, two different methods can be used: resampling with replacement (Yang, 2011; Nossent et al., 2011) or resampling without replacement (Pappenberger et al., 2008). When using resampling with replacement, the size of the resample data is the same as the size of the original sample. However, the bootstrap resample is most likely not identical to the original sample, because some data points might appear more than once in the resample, while others will not appear at all. For the resampling without replacement, a percentage of the original sample is randomly selected (e.g. 80%) and the original data points are not repeated in the resample data. Consequently, the size of the resample data is less than the size of the original sample and the values only appear once. In this research, we investigated the effect of different resampling methods on the estimation of the confidence intervals of Sobol' and PAWN sensitivity indices by applying and comparing the two techniques in the case of the NSE performance measure.

Considering the recommended values in the literature (Archer et al., 1997; Nossent et al., 2011; Pianosi and Wagener, 2015), the 95% confidence intervals are estimated using 1000 bootstrap resamples. For resampling without replacement, a random sub-set of 80% of the original samples is used (Pappenberger et al., 2008). Tests with different sub-set sizes provided almost similar results (not shown).

As shown in Figure B(a), resampling with replacement results in biased confidence intervals for the less-influential parameters (Slsubbsn, Smtmp, Timp). In fact, the PAWN indices of the less-influential parameters calculated over the original samples (gray bars in Figure B(a)) are systematically close to the lower bounds of the confidence intervals (red line) estimated by bootstrapping. Estimates of confidence intervals are instead considered unbiased if the results obtained over the original sample are bounded by the confidence intervals and close to the median values of those intervals, as for example happens for Slope in Figure B(a).

The reason for this bias in bootstrapping results when resampling with replacement is that the repeated presence of identical samples in the bootstrap resample causes small vertical "jumps" in the empirical distributions. For parameters with low PAWN indices, i.e. parameters with conditional and unconditional distributions close to each other, these small jumps may artificially increase the maximum vertical distance between the two distributions, which leads to overestimating the KS statistic (and hence the PAWN sensitivity index) when using bootstrap resamples. In contrast, resampling without replacement avoids repeating the same data points in the bootstrap resample, and does not induce biases (Figure B(b)).

In the Sobol' method, the bootstrapping results are unbiased, symmetric and median-centered for both resampling techniques (Figure B(c) and (d)). However, resampling with replacement (left panel) provides slightly wider confidence intervals, as compared to the resampling without replacement (right).

To conclude, our analysis shows that while for the Sobol' method there is no clear indication (at least from our case study application) that one resampling approach be better than another, for the PAWN method we find that bootstrapping with replacement provides biased estimates of confidence intervals and therefore should be avoided in favour of resampling without replacement, which provides unbiased results.

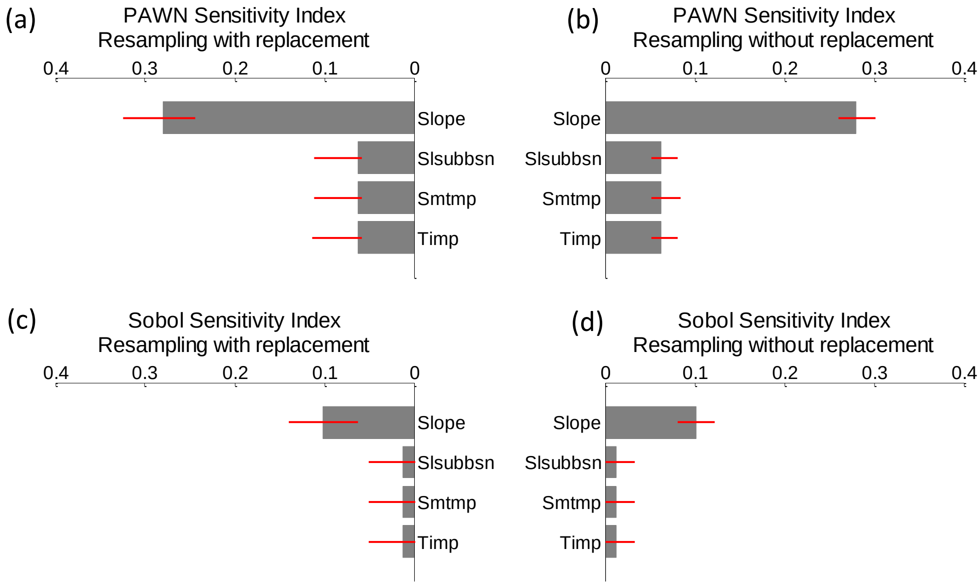

Figure B. For the PAWN indices, resampling with replacement provides biased 95% confidence intervals (red lines) for the parameters with low sensitivity indices (a), while almost unbiased results are obtained by using the resampling without replacement (b). For the Sobol' indices, both methods of resampling provide unbiased 95% confidence intervals. The grey bars represent the PAWN and Sobol' indices computed over the original samples.